

Mind on Statistics**Chapter 2****Sections 2.1 – 2.3**

1. Tallies and cross-tabulations are used to summarize which of these variable types?
 - A. Quantitative
 - B. Mathematical
 - C. Continuous
 - D. Categorical

KEY: D

2. The table below summarizes observed data on the gender and party membership of 1000 individuals:

Gender	Party Membership		
	Democrat	Republican	Total
Male	300	300	600
Female	200	200	400
Total	500	500	1000

Which one of the following statements about the relationship between gender and party is suggested by the data in the table?

- A. There is a relationship between gender and party membership.
- B. There is no relationship between gender and party membership.
- C. There is a relationship between gender and being a Democrat but not between gender and being a Republican.
- D. There is a relationship between gender and party membership for males but not for females.

KEY: B

3. Which one of these variables is a categorical variable?
 - A. Number of ear pierces a person has
 - B. Height of a person
 - C. Weight of a person
 - D. Opinion about legalization of marijuana

KEY: D

4. Which one of the following variables is not categorical?
 - A. Age of a person.
 - B. Gender of a person: male or female.
 - C. Choice on a test item: true or false.
 - D. Marital status of a person (single, married, divorced, other)

KEY: A

5. Which of the following is *not* a term used for a quantitative variable?
 - A. Measurement variable
 - B. Numerical variable
 - C. Continuous variable
 - D. Categorical variable

KEY: D

6. Listed below is the number of Congressional Medals of Honor awarded in wars fought by the United States.

War Medals	Civil	World War I	World War II	Korean	Vietnam	Other	Total
	1,520	124	440	141	239	105	2,569

What percent of all medals given were awarded during World War I and World War II?

- A. 4.83%
- B. 17.13%
- C. 21.95%
- D. 78.01%

KEY: C

Questions 7 to 9: In a survey of 1000 adults, respondents were asked about the expense of a college education and the relative necessity of financial assistance. The correspondents were classified as to whether they currently had a child in college or not (college status), and whether they thought the loan obligation for most college students was too high, about right, or too little (loan obligation opinion). The table below summarizes some of the survey results. Use these results to answer the following questions.

College Status	Loan Obligation Opinion			Total
	Too High	About Right	Too Little	
Child in College	350	80	10	440
No Child in College	250	200	110	560

7. Which type of variable is *Loan Obligation Opinion*?

- A. Categorical
- B. Quantitative
- C. Continuous
- D. Measurement

KEY: A

8. What role does the variable *Loan Obligation Opinion* play in this study?

- A. Explanatory
- B. Response
- C. Confounding
- D. It plays no role in the study.

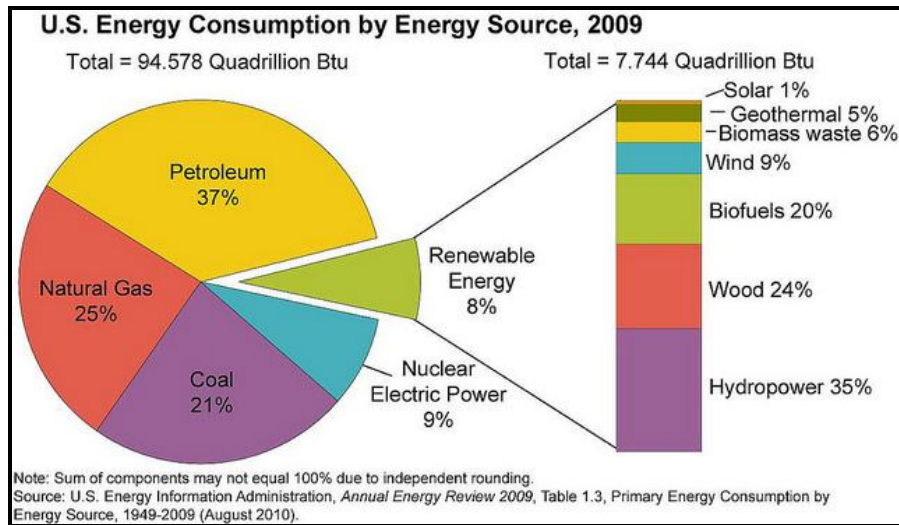
KEY: B

9. Which group had the greatest percentage of adults who thought loan obligations are too high?

- A. Those adults that have a child in college
- B. Those adults that do not have a child in college
- C. Both groups have the same percent thinking loan obligations are too high

KEY: A

Questions 10 to 13: The pie chart below shows the U.S. Energy Consumption by Energy Source for the year 2009.



10. The source with the highest consumption was

- A. petroleum.
- B. natural gas.
- C. coal.
- D. renewable energy.

KEY: A

11. The combined percent of petroleum and natural gas was

- A. less than 25% of the total energy consumption.
- B. between 25% and 50% of the total energy consumption.
- C. between 50% and 75% of the total energy consumption.
- D. more than 75% of the total energy consumption.

KEY: C

12. The consumption of natural gas in the United States totaled approximately how many quadrillion Btu?

- A. approximately 94.578 quadrillion Btu.
- B. approximately 25 quadrillion Btu.
- C. approximately 0.25 quadrillion Btu.
- D. approximately 23.6 quadrillion Btu.

KEY: D

13. The consumption of renewable sources in the United States totaled approximately 7.7 quadrillion Btu or about 8% of all energy used nationally. Over one-third of the consumption of renewable sources was from

- A. solar, geothermal, biomass waste, and wind combined.
- B. biofuels.
- C. wood.
- D. hydropower.

KEY: D

Chapter 2

14. Among 300 fatal car accidents, 135 were single-car crashes, 66 were two-car crashes, and 99 involved three or more cars. Calculate the relative frequency and percent of fatal car accidents by the number of cars involved.

KEY: Single car crashes 0.45 (45%); Two car crashes 0.22 (22%); Three car crashes 0.33 (33%).

15. The EPA sends out a survey to learn about people's water usage habits. Some of the questions included in the survey are given below.

Q1. How many times a week do you take a shower?

Q2. Do you leave the water running when you brush your teeth?

Q3. When you water your lawn, how long do you let the water run?

For each question, determine if it leads to categorical responses or quantitative responses.

KEY: Q1 and Q3 lead to quantitative responses, while Q2 leads to categorical responses.

Questions 16 and 17: A USA TODAY/CNN/Gallup Poll conducted April 19, 2005, was based on telephone interviews with 616 U.S. Catholics. One question asked was "When you think about your commitment to the Catholic Church, how much is your commitment affected by who the pope is -- a great deal, a moderate amount, not much, or not at all? The percentages are provided in the table below.

	Great deal	Moderate Amount	Not Much	Not at all	No Opinion
2005 APR 19	10%	32%	37%	20%	1%

16. The response variable being measured for this question could be called the *Commitment Status*. State what type of variable *commitment status* is and suggest an appropriate graph that could be made to display the distribution of this variable.

KEY: Commitment status is a categorical variable and an appropriate graph would be a pie chart or bar chart.

17. Approximately how many respondents stated that their commitment is not affected much?

KEY: 37% of 616 is 227.92 so approximately 228 respondents answered *not much*.

Section 2.4

18. The percent of data which lie between the lower and upper quartiles is
- A. 10%.
 - B. 25%.
 - C. 50%.
 - D. 75%.

KEY: C

19. A five-number summary for a data set is 35, 50, 60, 70, 90. About what percent of the observations are between 35 and 90?
- A. 25%
 - B. 50%
 - C. 95%
 - D. 100%

KEY: D

Questions 20 to 22: A five-number summary given in Case Study 1.1 for the fastest ever driving speeds reported by 102 women was: 30, 80, 89, 95, 130.

20. What is the interquartile range of these data?
- A. 6
 - B. 9
 - C. 15
 - D. 100

KEY: C

21. Fill in the blank in the following sentence: Approximately 25% of the women reported a fastest ever driving speed of at least _____ mph.
- A. 25
 - B. 80
 - C. 89
 - D. 95

KEY: C

22. Fill in the blank in the following sentence: Approximately 25% of the women reported a fastest ever driving speed of at most _____ mph.
- A. 30
 - B. 80
 - C. 89
 - D. 95

KEY: B

Questions 23 and 24: In a survey, students are asked how many hours they study in a typical week. A five-number summary of the responses is: 2, 9, 14, 20, 60.

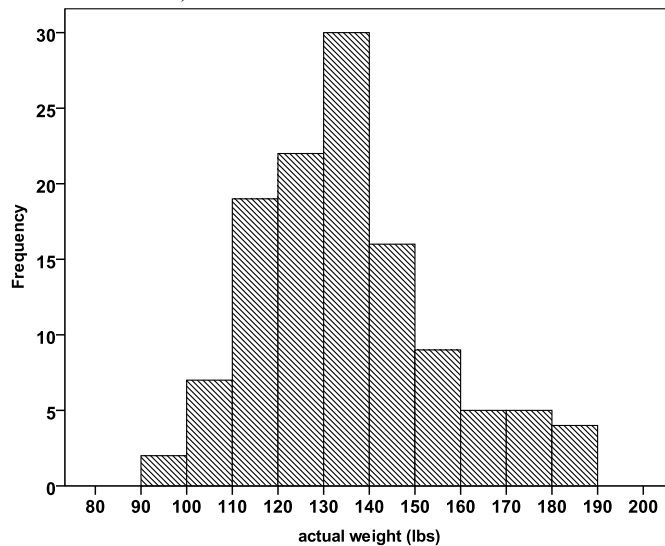
23. Which interval describes the number of hours spent studying in a typical week for about 50% of the students sampled?
- A. 2 to 9
 - B. 9 to 14
 - C. 9 to 20
 - D. 14 to 20

KEY: C

24. Fill in the blank in the following sentence. About 75% of the students spent at least ____ hours studying in a typical week.
- A. 9
 - B. 14
 - C. 20
 - D. 45

KEY: A

Questions 25 and 26: The following histogram is for the weights (lbs) of 119 female college students. (Source: *idealwtwomen* dataset on the CD.)



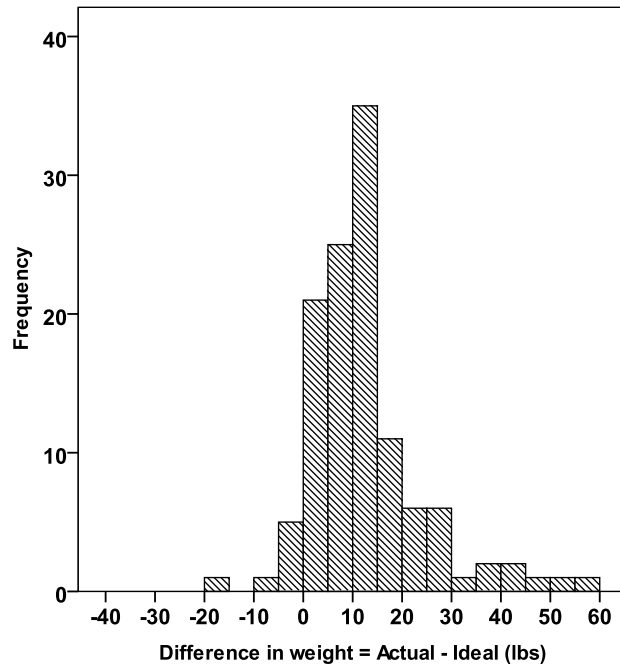
25. What is the approximate shape of the distribution?
- A. Nearly symmetric.
 - B. Skewed to the left.
 - C. Skewed to the right.
 - D. Bimodal (has more than one peak).

KEY: C

26. The best choice for the median weight for the 119 women based on the histogram is approximately
- A. 100 pounds.
 - B. 110 pounds.
 - C. 135 pounds.
 - D. 160 pounds.

KEY: C

Questions 27 to 29: The following histogram shows the distribution of the difference between the actual and “ideal” weights for 119 female college students. Ideal weights are responses to the question “What is your ideal weight”? The difference = actual – ideal. (Source: **idealwtwomen** dataset on the CD.)



27. What is the approximate shape of the distribution?
- Nearly symmetric.
 - Skewed to the left.
 - Skewed to the right.
 - Bimodal (has more than one peak).

KEY: C

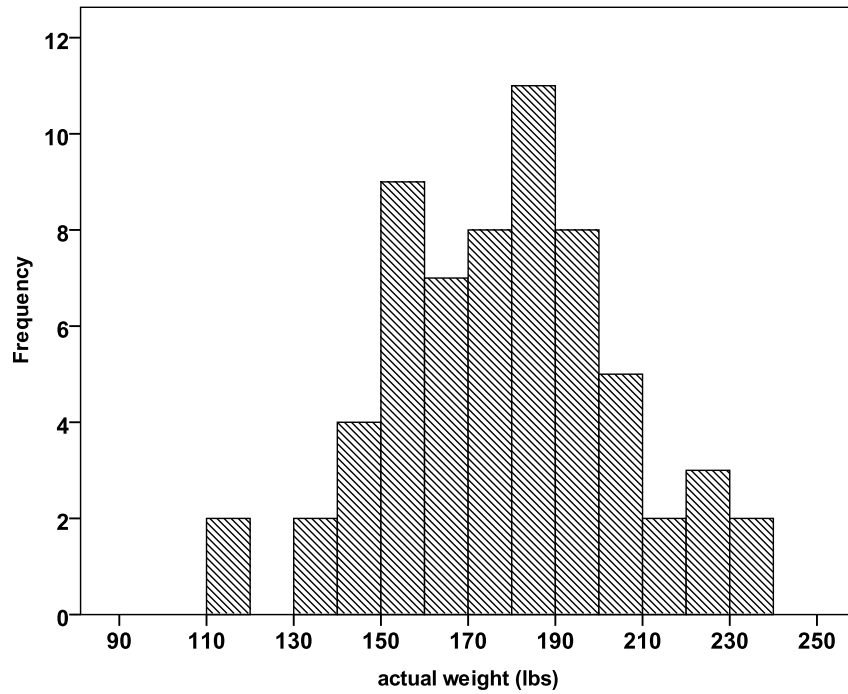
28. The median of the distribution is approximately
- 10 pounds.
 - 10 pounds.
 - 30 pounds.
 - 50 pounds.

KEY: B

29. Most of the women in this sample felt that their actual weight was
- about the same as their ideal weight.
 - less than their ideal weight.
 - greater than their ideal weight.
 - no more than 2 pounds different from their ideal weight.

KEY: C

Questions 29 and 30: The following histogram is for the weights (lbs) of 63 male college students. (Source: *idealwtmen* dataset on the CD.)



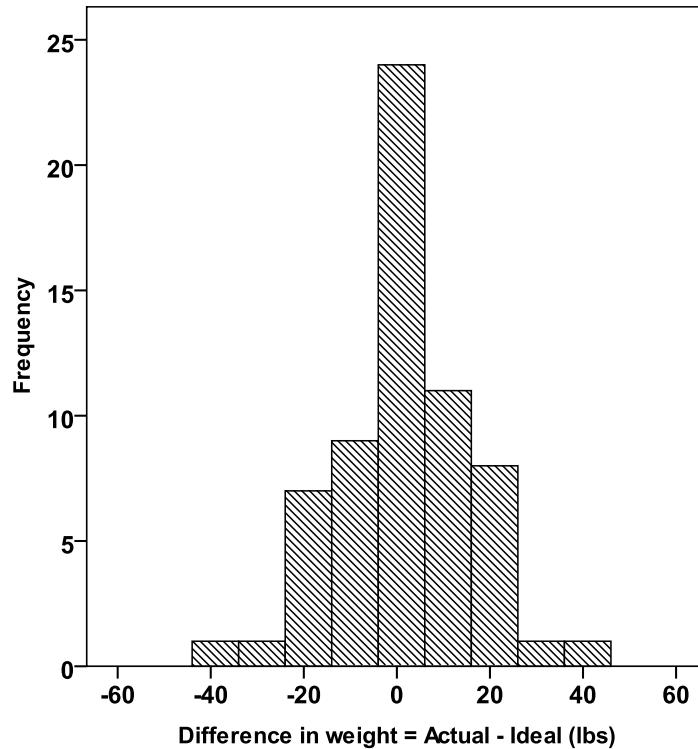
30. What is the best description for the approximate shape of this distribution?
- A. Nearly symmetric.
 - B. Skewed to the left.
 - C. Skewed to the right.
 - D. Bimodal (has more than one peak).

KEY: A

31. The median weight for the 63 men is approximately
- A. 130 pounds.
 - B. 150 pounds.
 - C. 180 pounds.
 - D. 220 pounds.

KEY: C

Questions 32 to 34: The following histogram gives the distribution of the difference between the actual and ideal weights for a sample of male college students. Ideal weights are responses to the question “What is your ideal weight”? The difference = actual – ideal.



32. What is the approximate shape of the distribution?
- Nearly symmetric.
 - Skewed to the left.
 - Skewed to the right.
 - Bimodal (has more than one peak).

KEY: A

33. The median of the distribution is approximately
- 10 pounds.
 - 0 pounds.
 - 10 pounds.
 - 20 pounds.

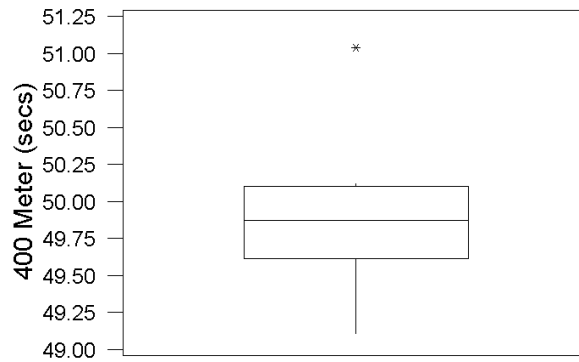
KEY: B

34. Most of the men in this sample felt that their actual weight was
- about the same as their ideal weight.
 - less than their ideal weight.
 - greater than their ideal weight.
 - no more than 2 pounds different from their ideal weight.

KEY: A

Chapter 2

35. The following boxplot is for the results of the women's 400-meter dash final race during the 2000 Olympics in Sydney, Australia. Cathy Freeman won in 49.11 seconds. Choose the correct statement about the boxplot.



- A. The median time is more than 50 seconds.
- B. The median time is less than 49.75 seconds.
- C. The fastest time of 49.11 seconds is an outlier.
- D. The slowest time of 51.04 seconds is an outlier.

KEY: D

36. Which of the following provides the most information about the shape of a data set?

- A. Boxplot
- B. Pie chart
- C. Five number summary
- D. Stem-and-leaf plot

KEY: D

Questions 37 to 39: According to a national sleep foundation survey, around 31 million Americans are sleep deprived. They also say women need more sleep than men and are being short-changed. Below are the five number summaries for the number of hours of sleep at night based on a survey of American men and women.

Men: 5.5, 6, 6.5, 7.5, 9

Women: 4.5, 5, 6, 7, 8

37. Write a sentence to compare men versus women in terms of the median amount of sleep at night

KEY: The survey shows that the median amount of sleep at night for women is 6 hours, about a half an hour less than that for men (which was 6.5 hours).

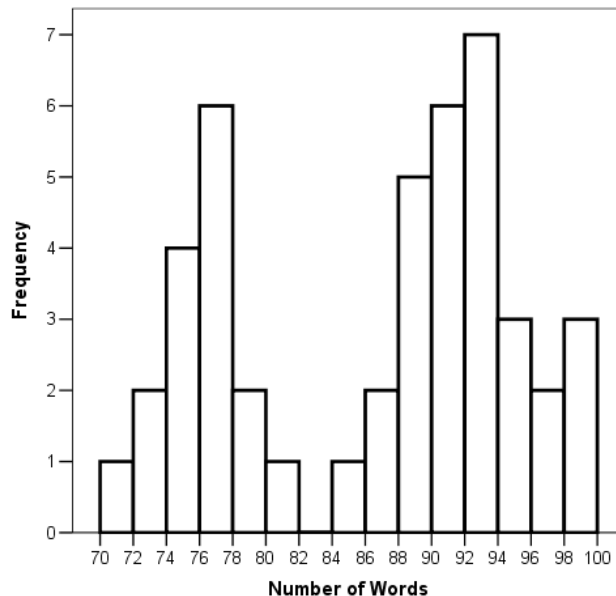
38. Write a sentence to compare men versus women in terms of the interquartile range for the amount of sleep at night.

KEY: Based on the survey, about 50% of the men get between 6 and 7.5 hours of sleep at night, while the interquartile range for women is from 5 to 7 hours of sleep.

39. What percent of women sleep at least 6 hours at night? What percent of men do so?

KEY: Based on the survey, about 50% of the women get at least 6 hours of sleep at night, while 75% of men do so.

Questions 40 to 42: A psychologist has developed a new technique intended to improve rote memory. To test the method against other standard methods, 45 high school students are selected at random and each is taught the new technique. The data on the number of words memorized correctly by the students were used to create the following histogram. Note the first class represents the interval [70, 72).



40. What proportion of students memorized correctly at least 94 words?

KEY: A total of 8 out of 45 students did so, for a proportion of 0.178 or 17.8%.

41. What is the overall shape of the distribution of the number of memorized words for these students?

KEY: The histogram is bimodal, thus showing evidence of two subgroups of students that perhaps should not be aggregated (or combined). The data should be further examined to try to identify the factor that has created these two subgroups.

42. Can we calculate the exact range of the 45 responses? If yes, calculate it. If no, explain why not.

KEY: No, we do not know the exact values for the minimum and maximum. We do know that the smallest observation is in the class [70, 72) and thus could be 70 or 71 words. Likewise the largest observation falls in the class [98, 100) and could be 98 or 99 words.

Section 2.5

43. What is the proper notation for the mean of a sample?

- A. \bar{x}
- B. μ
- C. σ
- D. s

KEY: A

44. A list of 5 pulse rates is: 70, 64, 80, 74, 92. What is the median for this list?

- A. 74
- B. 76
- C. 77
- D. 80

KEY: A

45. Which one of the following statements is most correct about a skewed dataset?

- A. The mean and median will usually be different.
- B. The mean and median will usually be the same.
- C. The mean will always be higher than the median.
- D. Whether the mean and median are the same depends on whether the data set is skewed to the right or to the left.

KEY: A

Questions 46 to 48: Listed below is a stem-and-leaf plot of the times it took 13 students to drink a 12 ounce beverage. Values for stems represent seconds and values for leaves represent tenths of a second.

```
3| 1234
3| 5
4| 0
5| 6
6| 11379
7|
8| 2
```

46. What was the median time to drink the beverage?

- A. 3.5 seconds.
- B. 4.0 seconds.
- C. 5.6 seconds.
- D. 6.9 seconds.

KEY: C

47. The lower quartile is

- A. 3.1 seconds.
- B. 3.35 seconds.
- C. 3.4 seconds.
- D. 3.5 seconds.

KEY: B

48. The upper quartile is

- A. 6.9 seconds.
- B. 6.5 seconds.
- C. 6.1 seconds.
- D. 5.6 seconds.

KEY: B

49. Which of the following would indicate that a dataset is skewed to the right?
- A. The interquartile range is larger than the range.
 - B. The range is larger than the interquartile range.
 - C. The mean is much larger than the median.
 - D. The mean is much smaller than the median.

KEY: C

50. If an exam was worth 100 points, and your score was at the 80th percentile, then
- A. your score was 80 out of 100.
 - B. 80% of the class had scores at or above your score.
 - C. 20% of the class had scores at or above your score.
 - D. 20% of the class had scores at or below your score.

KEY: C

Questions 51 to 53: The table below provides a statistical summary of the number of CDs owned as reported by students in a class survey done at Penn State University.

Variable	N	Mean	Minimum	Q1	Median	Q3	Maximum
CDs	250	85	0	30	50	100	500

51. Approximately what percent of students own somewhere between 30 and 50 CDs?
- A. 50%
 - B. 25%
 - C. 20%
 - D. 4%

KEY: B

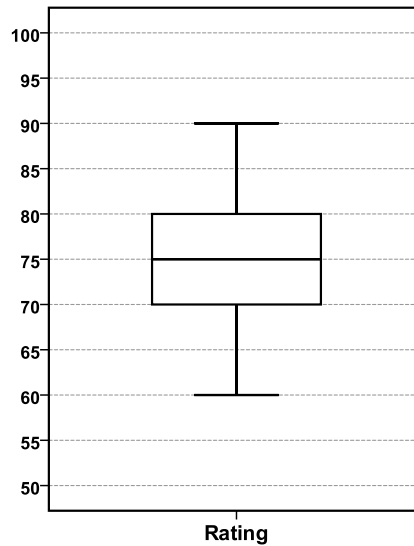
52. What is the interquartile range for these data?
- A. 500
 - B. 100
 - C. 70
 - D. 30

KEY: C

53. Based on the summary shown, which of the following statements most likely describes the shape of the CDs owned dataset?
- A. The summary is evidence that the data are symmetric and bell-shaped.
 - B. The summary is evidence that the data are symmetric but not bell-shaped.
 - C. The summary is evidence that the data are skewed to the left.
 - D. The summary is evidence that the data are skewed to the right.

KEY: D

Questions 54 to 57: The following boxplot gives the distribution of the ratings of a new brand of peanut butter for 50 randomly selected consumers (100 points possible with higher points corresponding to a more favorable rating).



54. Based on the boxplot,
- A. the distribution appears to be skewed to the left.
 - B. the distribution appears to be skewed to the right.
 - C. there appear to be outliers at about 60 and 90.
 - D. there do not appear to be any outliers.

KEY: D

55. The lower quartile is between
- A. 50 to 60 points.
 - B. 60 to 70 points.
 - C. 70 to 75 points.
 - D. 80 to 90 points.

KEY: B

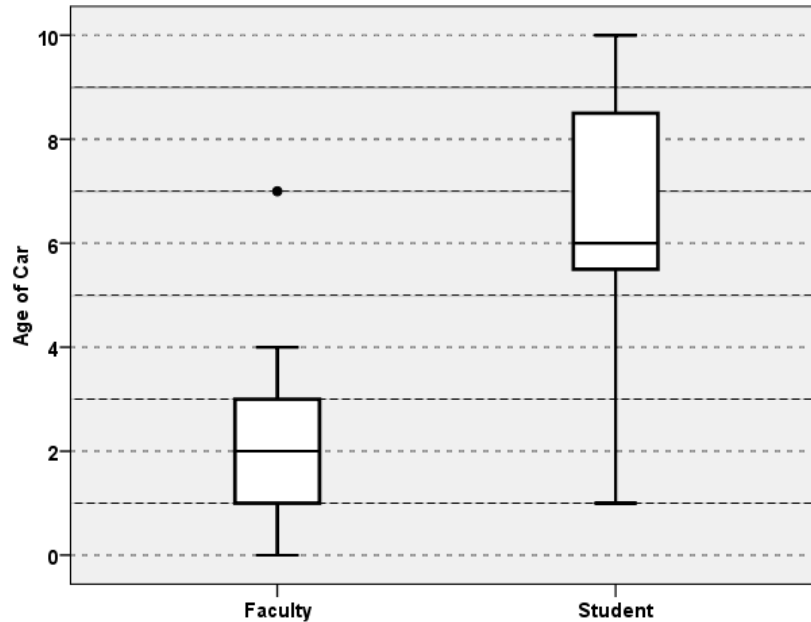
56. The median is
- A. 60 points.
 - B. 70 points.
 - C. 75 points.
 - D. 80 points.

KEY: C

57. The upper quartile is between
- A. 50 to 60 points.
 - B. 60 to 70 points.
 - C. 70 to 75 points.
 - D. 80 to 90 points.

KEY: D

58. A recent study was conducted to compare the age of vehicles in a student parking lot versus those in a faculty parking lot at a major university. A random sample of 15 cars was taken from each lot and the age of the car was recorded by taking the current year and subtracting the model year from it. The two boxplots are shown below to summarize these results.



Compare the two distributions based on these side-by-side boxplots.

- KEY:** Overall the faculty cars are newer with a median age of 3 years compared to a median age of 6 years for students' cars. There was (at least) one faculty car that was much older than the rest and thus shown as an outlier at 7 years. The overall range for the ages of the students' cars is larger ($10 - 1 = 9$ years) as compared to the range for the ages of the cars owned by faculty ($7 - 0 = 7$ years).

Section 2.6

59. An outlier is a data value that
- A. is larger than 1 million.
 - B. equals the minimum value in a set of data.
 - C. equals the maximum value in a set of data.
 - D. is not consistent with the bulk of the data.

KEY: D

60. Which statistic is *not* resistant to an outlier in the data?
- A. Lower quartile
 - B. Upper quartile
 - C. Median
 - D. Mean

KEY: D

61. Which one of these statistics is unaffected by outliers?
- A. Interquartile range
 - B. Mean
 - C. Standard deviation
 - D. Range

KEY A

62. Which one of the following statistics would be affected by an outlier?
- A. Median
 - B. Standard deviation
 - C. Lower quartile
 - D. Upper quartile

KEY: B

63. Which of the following could account for an outlier in a dataset?
- A. Natural variability in the measurement of interest.
 - B. Recording the wrong category for an individual's value of a categorical variable.
 - C. A symmetric distribution for the measurement of interest.
 - D. Measuring more than one variable for each individual.

KEY: A

64. Determine whether the following statement is true or false and explain your answer: Outliers cause complications in all statistical analyses.
KEY: False, outliers do affect some statistics such as means and standard deviations. However, there are appropriate measures of location and spread if outliers are present and cannot be discarded, namely, the median and the interquartile range.
65. Determine whether the following statement is true or false and explain your answer: Since outliers cause complications in statistical analyses, they should be discarded before computing summaries such as the mean and the standard deviation.
KEY: False. Although outliers do affect summaries such as the mean and standard deviation, they should never be discarded without justification.
66. What is a reasonable action if an outlier is a legitimate data value and represents natural variability for the group and variable measured?
KEY: The value should not be discarded; in fact, it may be one of the more interesting values in the data set.
67. What is a reasonable action if an outlier was a mistake made in measuring the object?
KEY: The value should be corrected if possible or discarded if not possible to correct it.
68. What is a reasonable action if an outlier is the value for the only young subject in a sample where all of the other values were for older subjects?
KEY: The value should be discarded and the results summarized and reported for the older subjects only.

Section 2.7

69. Which choice lists two statistics that give information only about the location of a dataset and not the spread?
- A. IQR and standard deviation
 - B. Mean and standard deviation
 - C. Median and range
 - D. Mean and median

KEY: D

70. Which of the following measures is not a measure of spread?
- A. Variance
 - B. Standard deviation
 - C. Interquartile range
 - D. Median

KEY: D

71. Which one of the following summary statistics is *not* a measure of the variation (spread) in a data set?
- A. Median
 - B. Standard deviation
 - C. Range
 - D. Interquartile range

KEY: A

72. The head circumference (in centimeters) of 15 college-age males was obtained, resulting in the following measurements: 55, 56, 56, 56.5, 57, 57, 57, 57.5, 58, 58, 58, 58.5, 59, 59, 63. If the last measurement (63 cm's) were incorrectly recorded as 73, which one of the following statistics would change?
- A. Q1 (1st quartile)
 - B. Standard deviation
 - C. Median
 - D. Q3 (3rd quartile)

KEY: B

73. Which of the following is true about the relationship between the standard deviation s and the range for a large bell-shaped data set?
- A. The range is approximately $1/2$ of a standard deviation.
 - B. The range is approximately 2 standard deviations.
 - C. The range is approximately 6 standard deviations.
 - D. The range is approximately $1/6$ of a standard deviation

KEY: C

74. By inspection, determine which of the following sets of numbers has the smallest standard deviation.
- A. 2, 3, 4, 5
 - B. 4, 4, 4, 5
 - C. 0, 0, 5, 5
 - D. 5, 5, 5, 5

KEY: D

75. The mean hours of sleep that students get per night is 7 hours, the standard deviation of hours of sleep is 1.7 hours, and the distribution is approximately normal. Complete the following sentence. For about 95% of students, nightly amount of sleep is between _____.
- A. 5.3 and 8.7 hrs
 - B. 5 and 9 hrs
 - C. 3.6 and 10.4 hrs
 - D. 1.9 and 12.1 hrs

KEY: C

76. For a large sample of blood pressure values, the mean is 120 and the standard deviation is 10. Assuming a bell-shaped curve, which interval is likely to be about the interval that contains 95% of the blood pressures in the sample?
- A. 110 to 130
 - B. 100 to 140
 - C. 90 to 150
 - D. 50 to 190

KEY: B

77. For a large sample of blood pressure values, the mean is 120 and the standard deviation is 10. Assuming a bell-shaped curve, which interval is likely to be about the interval from the minimum to maximum blood pressures in the sample?
- A. 120 to 150
 - B. 110 to 130
 - C. 90 to 150
 - D. 50 to 190

KEY: C

78. Which of the following would indicate that a dataset is *not* bell-shaped?
- A. The range is equal to 5 standard deviations.
 - B. The range is larger than the interquartile range.
 - C. The mean is much smaller than the median.
 - D. There are no outliers.

KEY: C

79. The possible values for a standardized score (z -score)
- A. can be any number: positive, negative, or 0.
 - B. must be within the range from -3 to 3
 - C. must be non-negative.
 - D. must be strictly positive.

KEY: A

80. Which of the following best describes the standardized (z) score for an observation?
- A. It is the number of standard deviations the observation falls from the mean.
 - B. It is the most common score for that type of observation.
 - C. It is one standard deviation more than the observation.
 - D. It is the center of the list of scores from which the observation was taken.

KEY: A

Chapter 2

81. Scores on an achievement test averaged 70 with a standard deviation of 10. Serena's score was 85. What was her standardized score (also called a z -score)?
- A. -1.5
 - B. 1.5
 - C. 15
 - D. 85

KEY: B

Questions 82 to 85: Suppose that amount spent by students on textbooks this semester has approximately a bell-shaped distribution. The mean amount spent was \$300 and the standard deviation is \$100.

82. Which choice best completes the following sentence? About 68% of students spent between ____.
- A. \$300 and \$400
 - B. \$200 and \$400
 - C. \$100 and \$500
 - D. \$266 and \$334

KEY: B

83. What amount spent on textbooks has a standardized score equal to 0.5?
- A. \$150
 - B. \$250
 - C. \$300.50
 - D. \$350

KEY: D

84. What percent of students spent more than \$350?
- A. 50%
 - B. 0.5%
 - C. 69.15%
 - D. 30.85%

KEY: D

85. A student spent \$500 on textbooks. What percentile does their value correspond to?
- A. 97.5th percentile
 - B. 95th percentile
 - C. 5th percentile
 - D. 2.5th percentile

KEY: A

86. Explain the difference between the population standard deviation σ and the sample standard deviation s .
KEY: The population standard deviation is a measure of spread in the population and is a parameter (fixed value, usually unknown). The sample standard deviation is an estimate of the population standard deviation and is a statistic.

87. For each of the following numerical summaries, decide whether it is a resistant statistic or not: mean, median, standard deviation, range, interquartile range.

KEY: Resistant statistics would include the median and the interquartile range. Non-resistant statistics would include the mean, the standard deviation, and the range.

88. Suppose that the average height for college men is 66 inches. If the height distribution is bell-shaped, and 95% of the men have heights between 60 inches and 72 inches, what is the standard deviation of heights for this population?

KEY: 3 inches

Questions 89 to 91: The average rainfall during the month of November in San Francisco, California, is 2.62 inches. The standard deviation is 2.79 inches.

89. What is the standardized score (z -score) for 5.18 inches, the rainfall in San Francisco during November 2001?

KEY: 0.918

90. What is the standardized score (z -score) for 11.78 inches, the rainfall in San Francisco during November 1885?

KEY: 3.28

91. What is the standardized score (z -score) for 1 inch of rain in November?

KEY: -0.581

Questions 92 to 94: Suppose that the average number of years to graduate at a university is 4 years, with a standard deviation of 0.5 years. Assume a bell-shaped distribution for years to graduate.

92. From the Empirical Rule, what is a range of values that 68% of the students should graduate between?

KEY: 3.5 to 4.5 years

93. From the Empirical Rule, what is a range of values that 95% of the students should graduate between?

KEY: 3 to 5 years

94. From the Empirical Rule, what is a range of values that 99.7% of the students should graduate between?

KEY: 2.5 to 5.5 years

